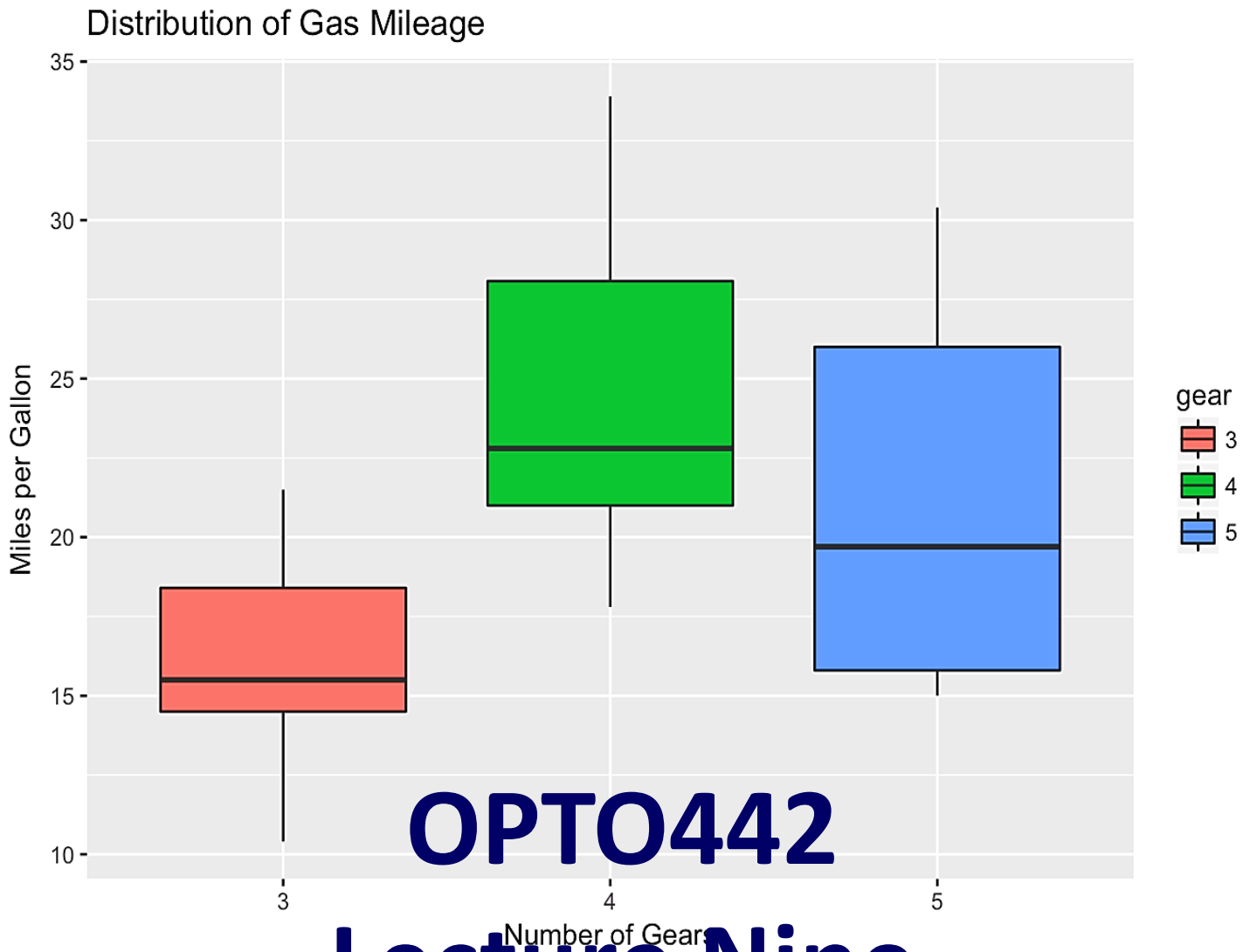
جامعة الملك سعود
King Saud University

OPTO571

Experimental Design and Data Management

GAMAL EL-HITI

• Side By Side Boxplots



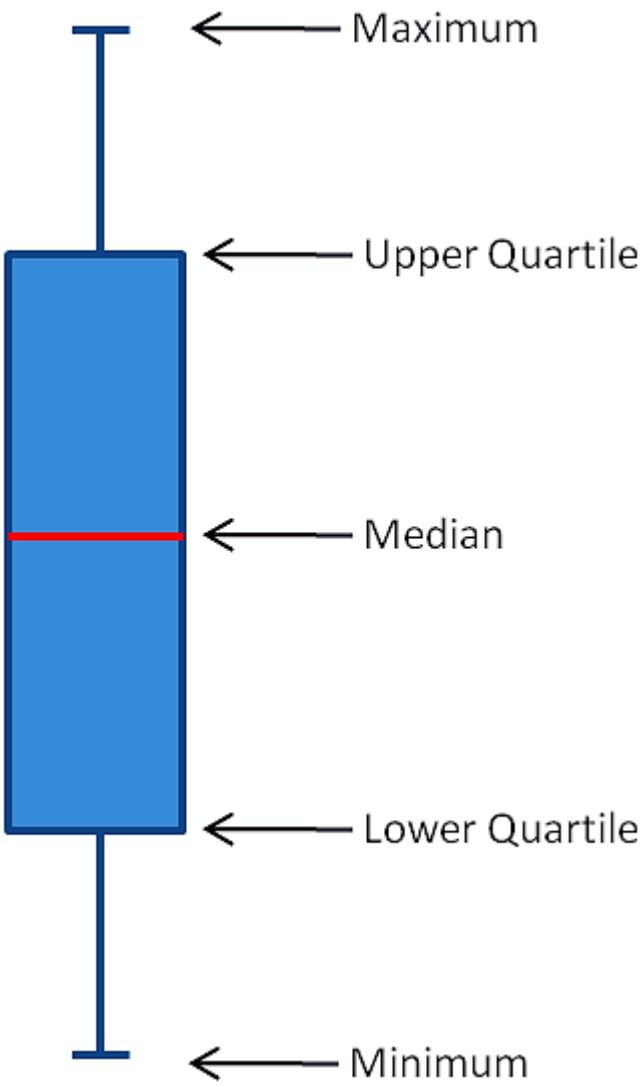
OPTO442

Lecture Nine

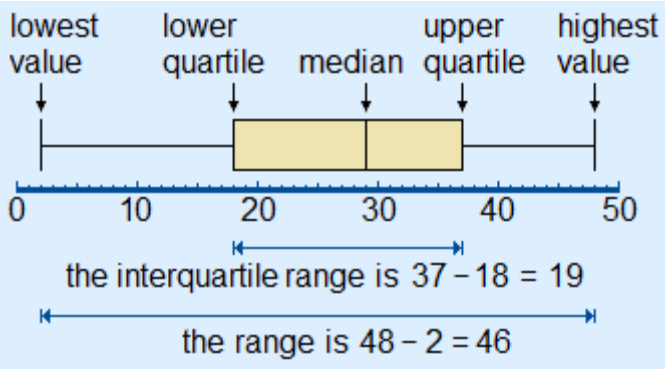
Side by Side Boxplots

- Graphical Investigation
- Side-by-side boxplot is one of the best method to differentiate between groups of data graphically.
- Whether the differences between the groups are significant depends on:
  - The difference in the means
  - The standard deviations of each group
  - The sample sizes

Side by Side Boxplots

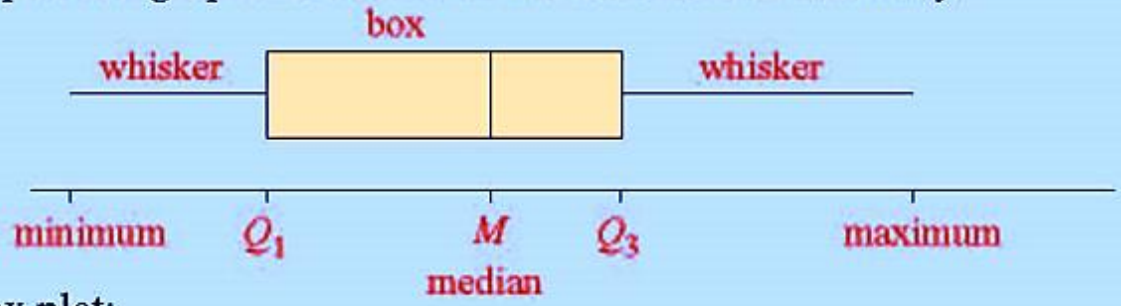


Side by Side Boxplots



The box plot

A box plot is a graphical version of the five-number summary.



In a box plot:

- a box is used to represent the middle 50% of scores
- the median is shown by a vertical line drawn within the box
- lines (called whiskers) are extended out from the lower and upper ends of the box to the smallest and largest data values of the data set respectively

Side by Side Boxplots

- Using a Stacked 2 Column Chart
- The order in which we prepare the five numbers statistics for this type of Boxplots is different then normal one.
- The order is:
- MAXIMUM
- Q1
- MEDIAN
- Q3
- MINIMUM

Side by Side Boxplots

Subjects: 25 patients with eye infection.

Treatment: Treatment A, Treatment B, and Treatment C.

Measurement: No of days until infection heal.

Data [and means]:

A: 5, 6, 6, 7, 7, 8, 9, 10 [7.25]

B: 7, 7, 8, 9, 9, 10, 10, 11 [8.875]

C: 7, 9, 9, 10, 10, 10, 11, 12, 13 [10.11]

- Draw side by side boxplots?

Side by Side Boxplots

Values	Treatment A	Treatment B	Treatment C
1	5	7	7
2	6	7	9
3	6	8	9
4	7	9	10
5	7	9	10
6	8	10	10
7	9	10	11
8	10	11	12
9			13
Statistic	A	B	C
MAX	10	11	13
Q1	6	7.75	9
MEDIAN	7	9	10
Q3	8.25	10	11
MIN	5	7	7

Side by Side Boxplots

- Since we will be using a stacked 2D Column Chart we have to modify our data.
- These changes are only made to enable us to draw the chart properly.
- MAXIMUM value (that is the length of the top whisker) needs to be changed.
- Q1 or 1st QUARTILE remains unchanged.
- MEDIAN needs to be changed.
- Q3 or 3rd QUARTILE needs to be changed.
- MINIMUM needs to be changed.

Side by Side Boxplots

- New MAXIMUM = MAX – Q3

  1.75, 1 and 2

- New Q1 is the same as old Q1

  6, 7.75 and 9

- New MEDIAN = MEDIAN - Q1

  1, 1.25 and 1

- New Q3 = Q3 – MEDIAN

  1.25, 1 and 1

- New MINIMUM = Q1 – MIN

  1, 0.75 and 2

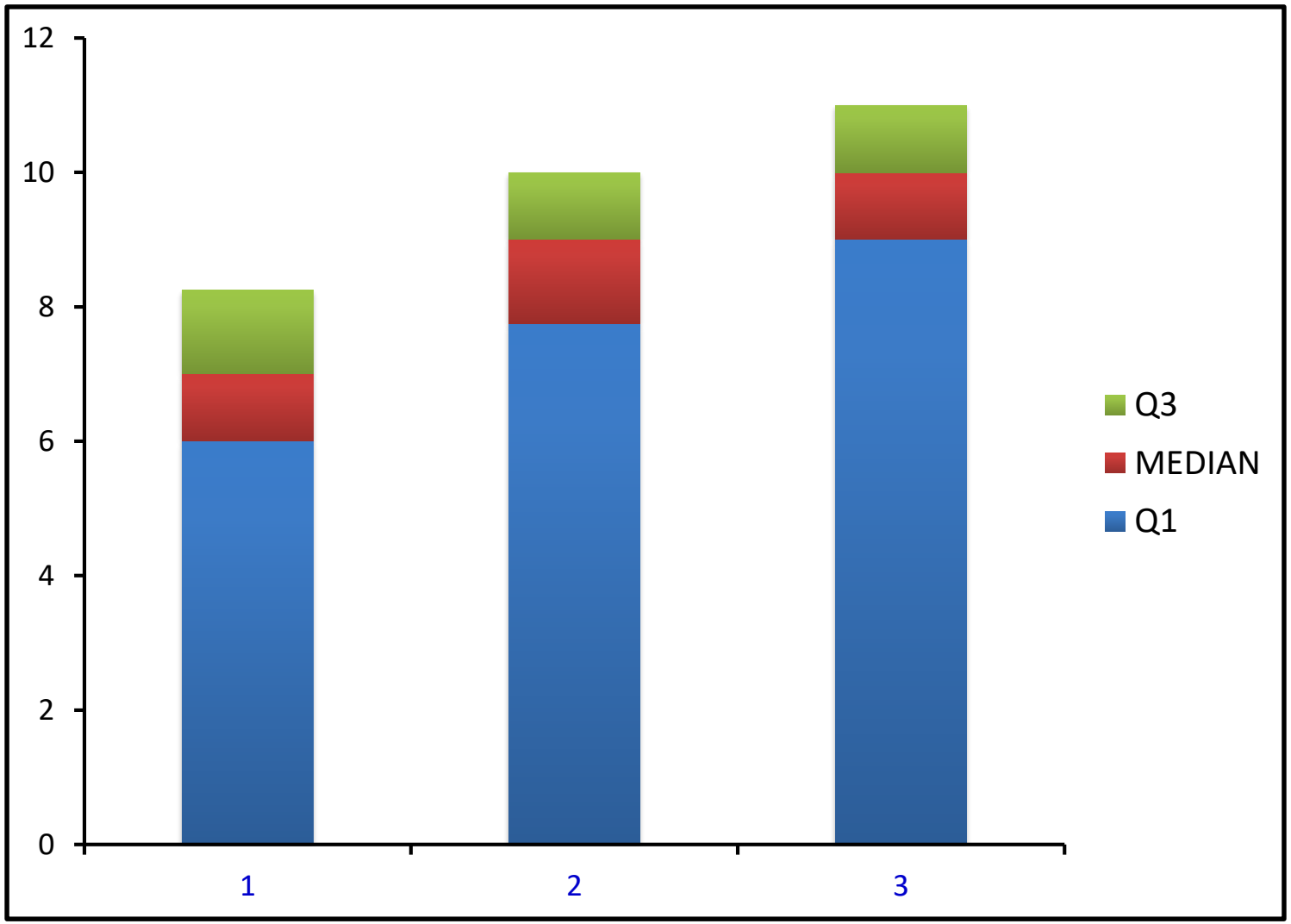
Side by Side Boxplots

Statistic	Treatment A	Treatment B	Treatment C
MAX	10	11	13
Q1	6	7.75	9
MEDIAN	7	9	10
Q3	8.25	10	11
MIN	5	7	7
New Statistic	A	B	C
MAX	1.75	1	2
Q1	6	7.75	9
MEDIAN	1	1.25	1
Q3	1.25	1	1
MIN	1	0.75	2

Side by Side Boxplots

- Select the title, Q3, MEDIAN and Q1 results only.
- Then select the following in order:
- Insert
- Charts
- Column
- 2D Column
- Stacked

Side by Side Boxplots



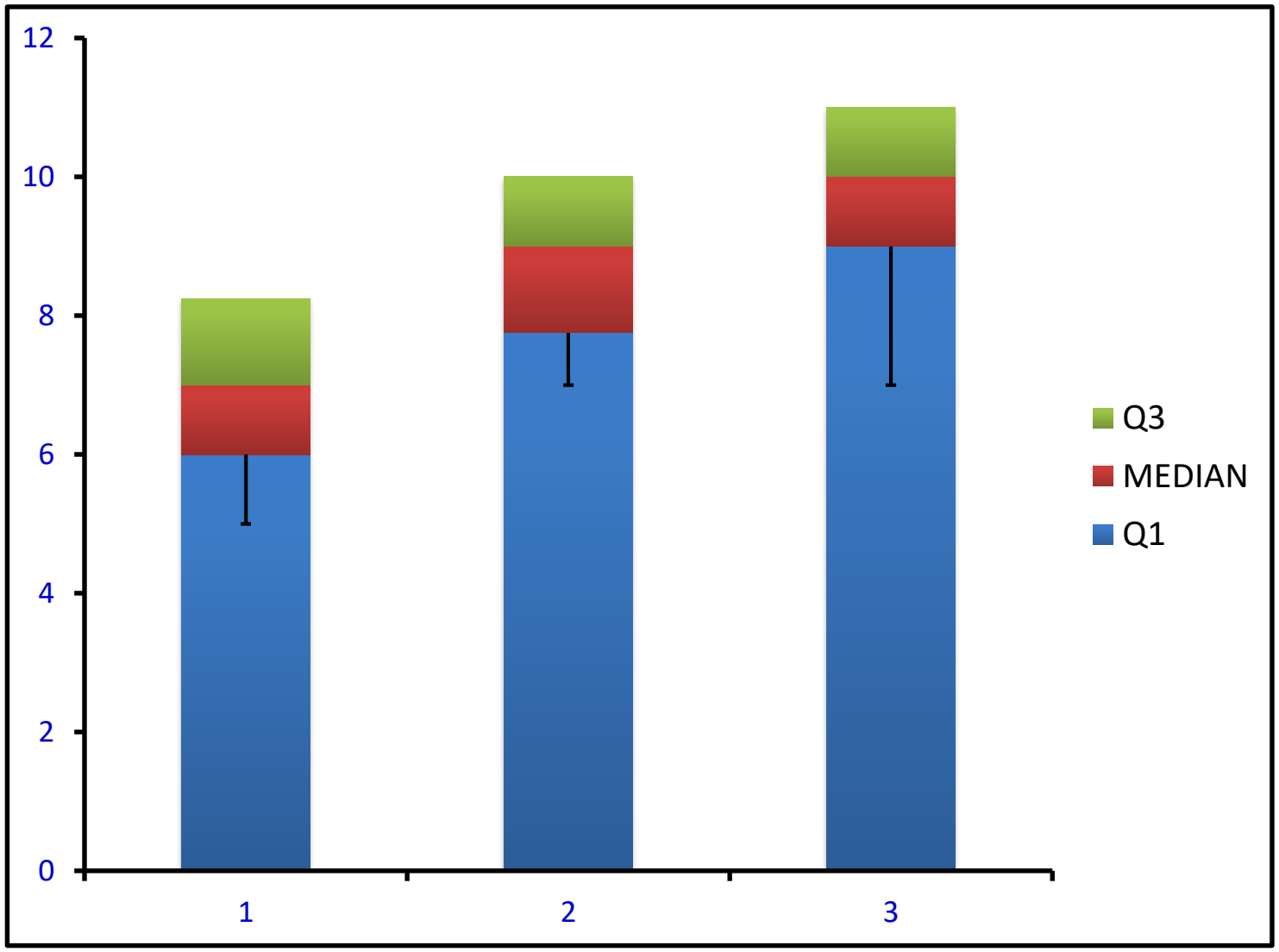
Side by Side Boxplots

- Add the minimum value whiskers.
- Select the Q1 bar on the chart.
- Chart Tools , Layout (or Add Chart Element) and select Error Bars.
- More Error Bars options.
- Select the Display Direction: Minus.
- Indicate the Error Amount: Custom.
- Click the Specify Value button.
- Leave the Positive Error Value as it is.
- Select the MIN values in the data set as the Negative Error Value.

Side by Side Boxplots

- Or select the Q1 bar on the chart.
- Select Error Bars.
- More Error Bars options
- Select the Display Direction: Minus.
- Indicate the Error Amount: Custom.
- Click the Specify Value button.
- Leave the Positive Error Value as it is.
- Select the MIN values in the data set as the Negative Error Value.

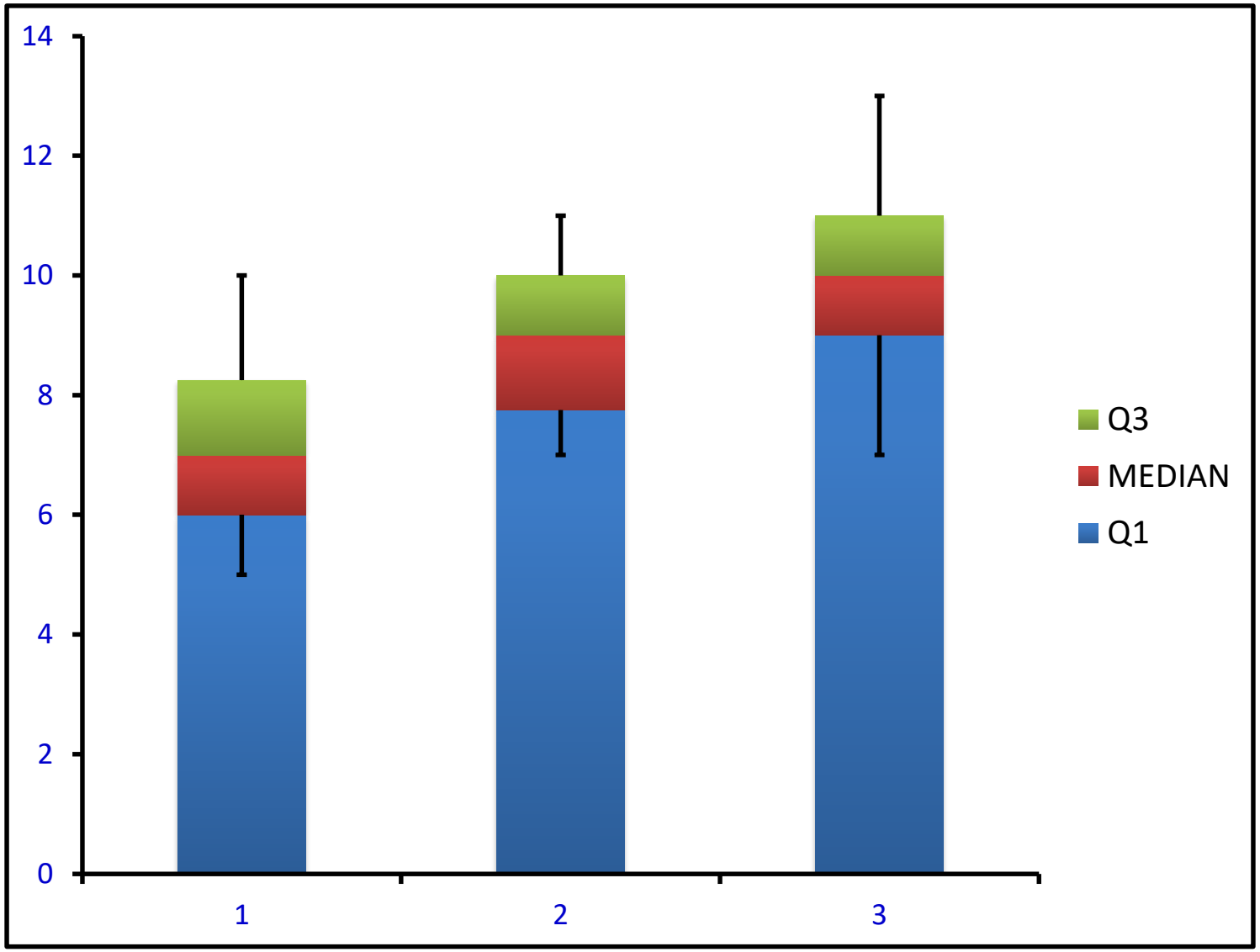
Side by Side Boxplots



Side by Side Boxplots

- For the max value whiskers begin by selecting the Q3 bar on the chart.
- More Error Bars options.
- Select the Display Direction: Plus.
- Indicate the Error Amount: Custom.
- Click the Specify Value button.
- Leave the Negative Error Value as it is.
- Select the MAX values in the data set as the Positive Error Bar.

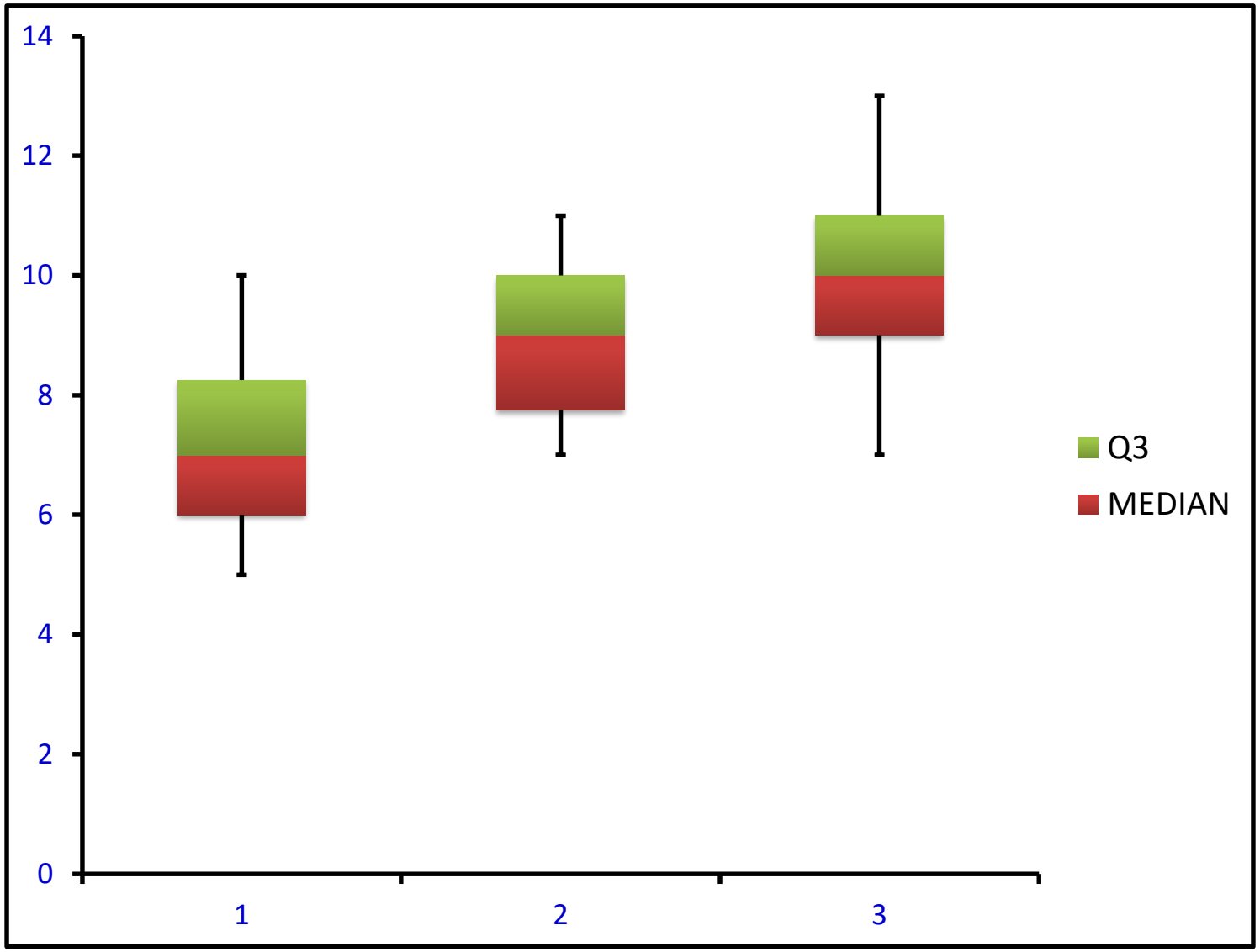
Side by Side Boxplots



Side by Side Boxplots

- Make the Q1 bars invisible now:
- Click on any of the Q1 Bars: any of them and they will all be highlighted
- Right Hand Click
- Format Data Series
- Fill
- No Fill
- Border Color
- No Line

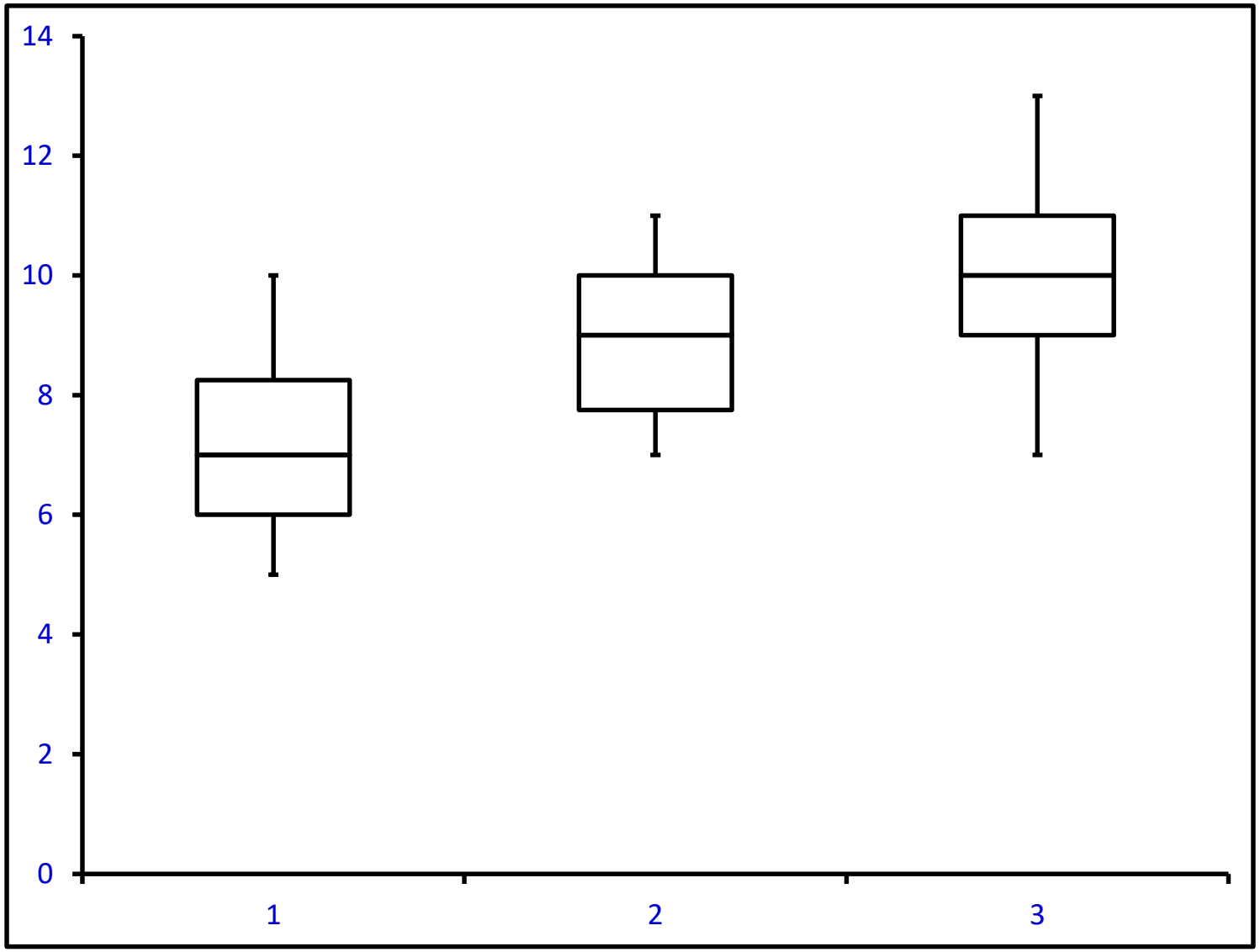
Side by Side Boxplots



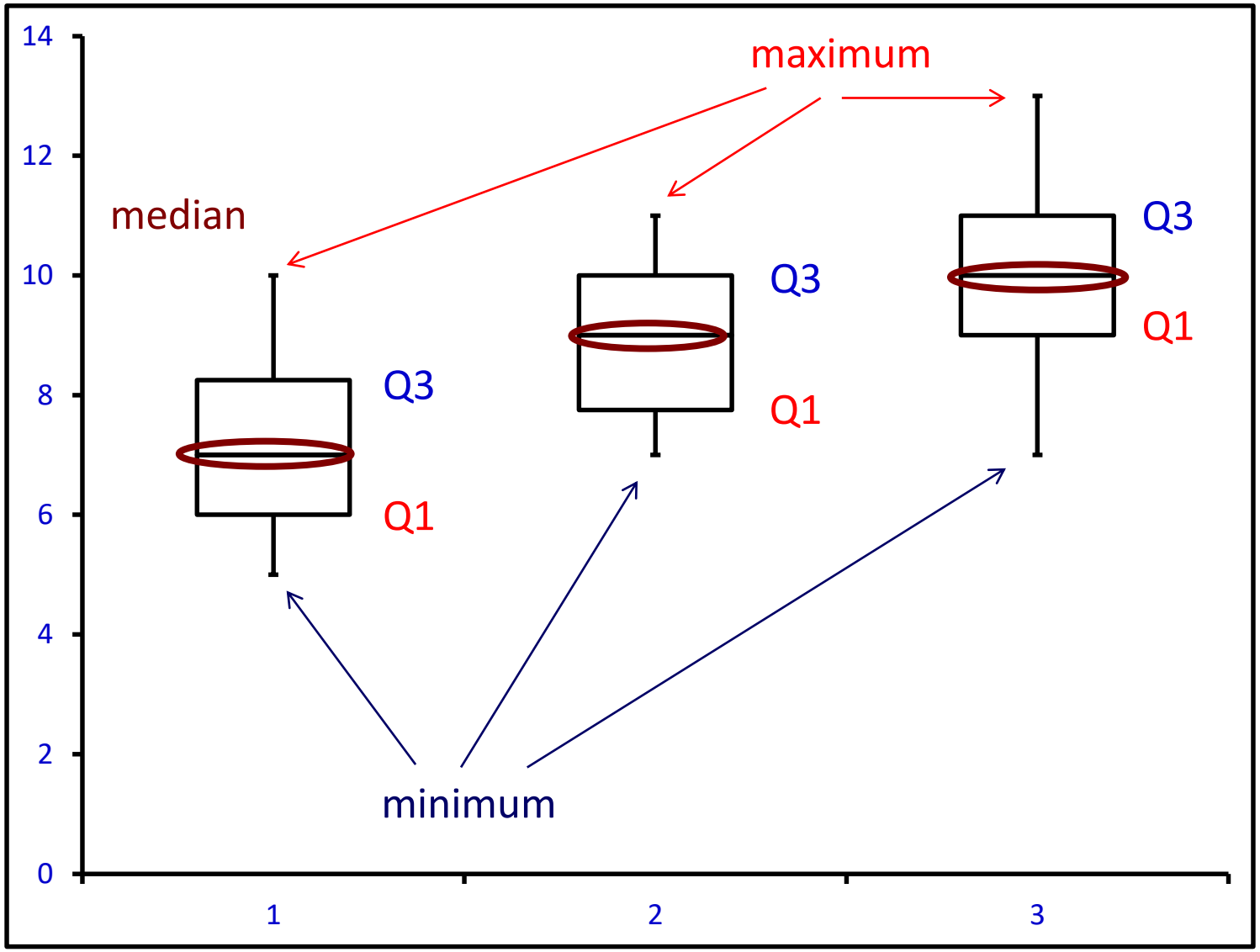
Side by Side Boxplots

- For the remaining two data series, Q3 and MEDIAN, format them to have:
- No Fill
- Border Colour
- Solid Line
- Color
- Black
- Delete the Legend

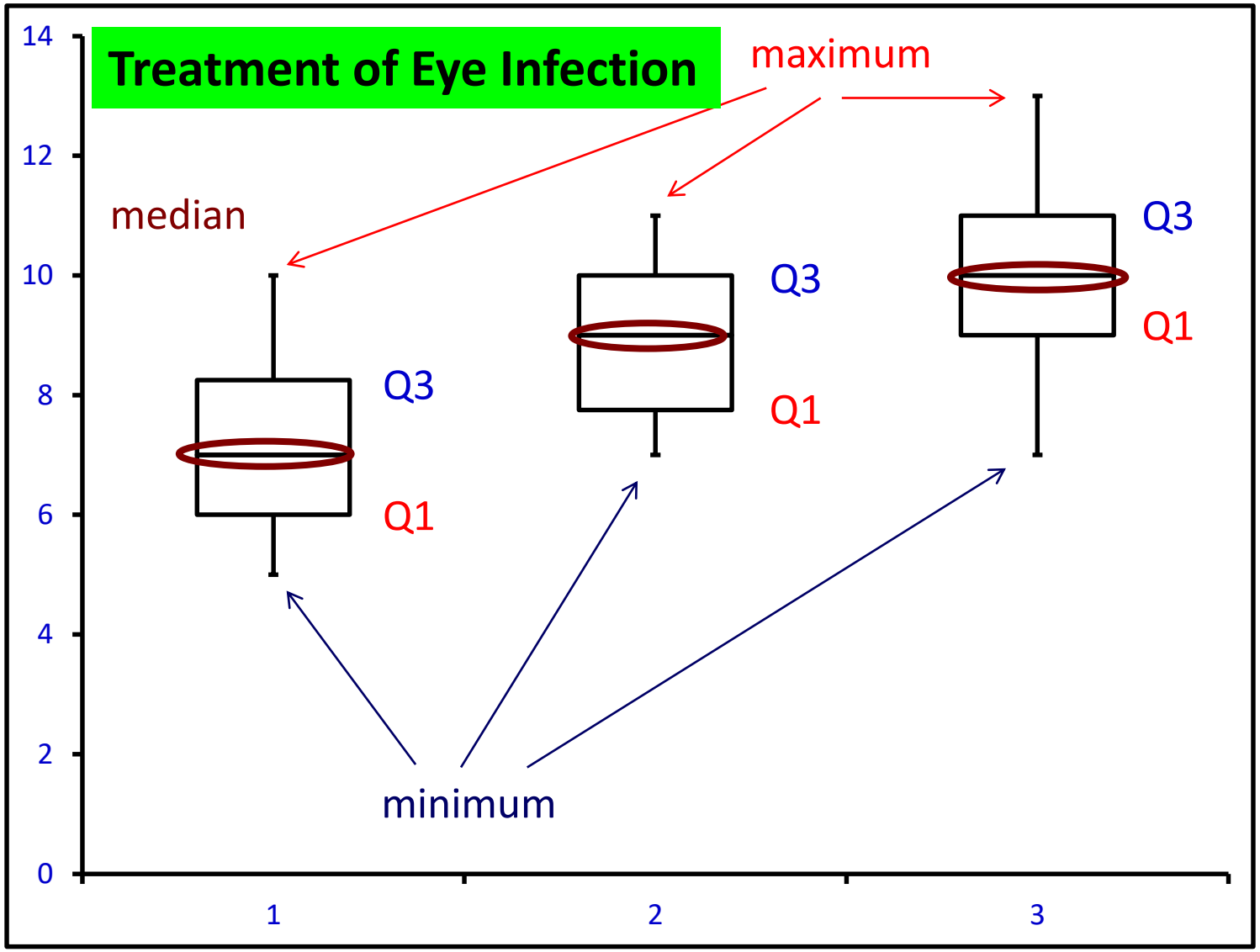
Side by Side Boxplots



Side by Side Boxplots

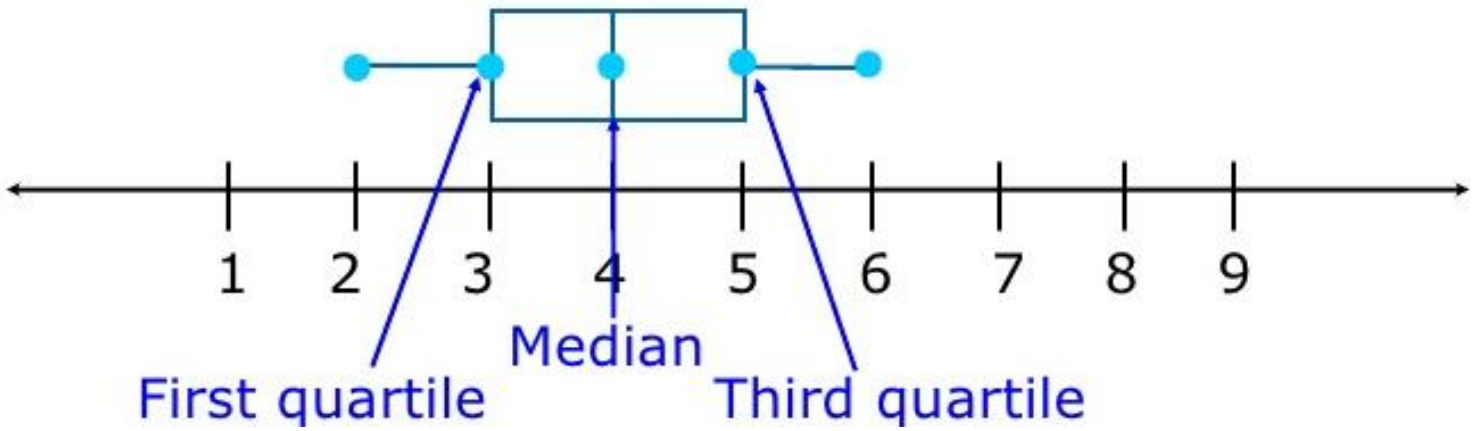


Side by Side Boxplots



Side by Side Boxplots

A **box-and-whisker plot** shows the distribution of data. The middle half of the data is represented by a "box" with a vertical line at the median. The lower fourth and upper fourth quarters are represented by "whiskers" that extend to the smallest and largest values.



Side by Side Boxplots

Use the given data to make a box-and-whisker plot:
21, 25, 15, 13, 17, 19, 19, 21

Step 1. Order the data and find the smallest value, first quartile, median, third quartile, and largest value.

13 15 17 19 19 21 21 25

smallest value: 13 largest value: 25

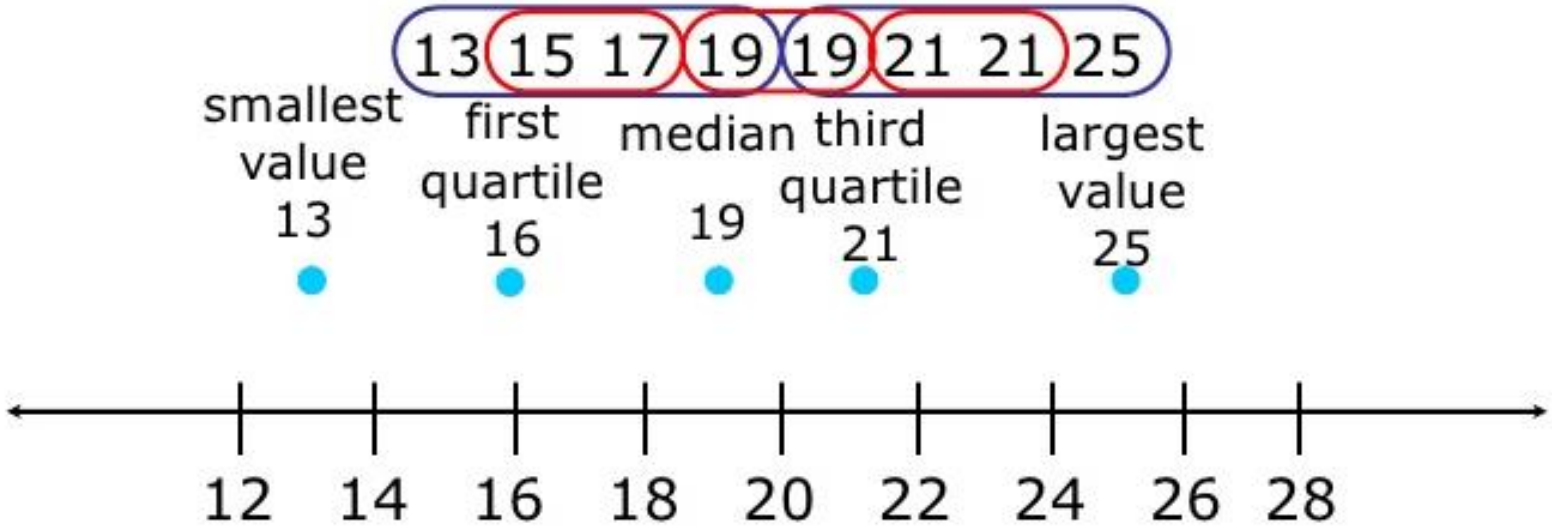
first quartile: $\frac{15 + 17}{2} = 16$ third quartile: $\frac{21 + 21}{2} = 21$

median: $\frac{19 + 19}{2} = 19$

Side by Side Boxplots

Use the given data to make a box-and-whisker plot.

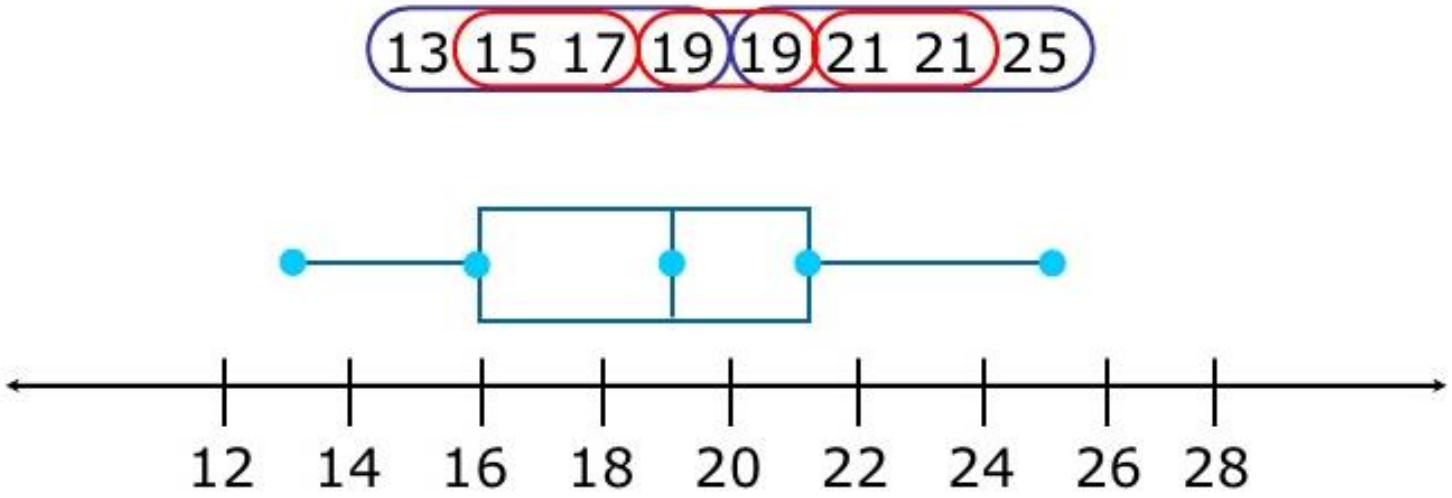
Step 2. Draw a number line and plot a point above each value from Step 1.



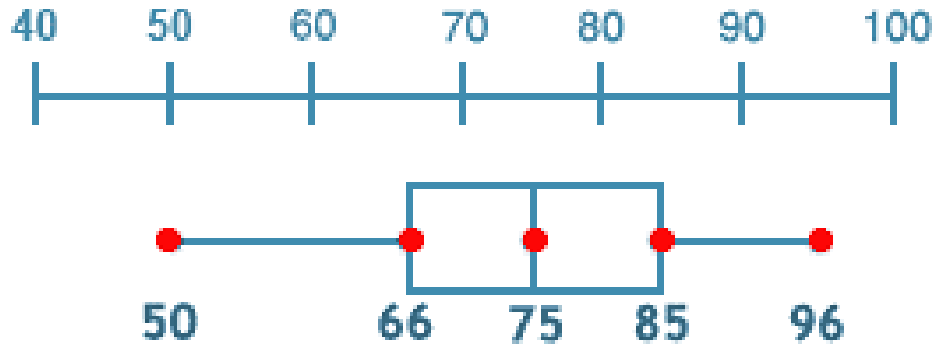
Side by Side Boxplots

Use the given data to make a box-and-whisker plot.

Step 3. Draw the box and whiskers. The left whisker is from the 1st to 2nd dot. The right whisker connects the 4th and 5th dot. The box connects the 2nd dot and 4th dot. Draw a line through the median.

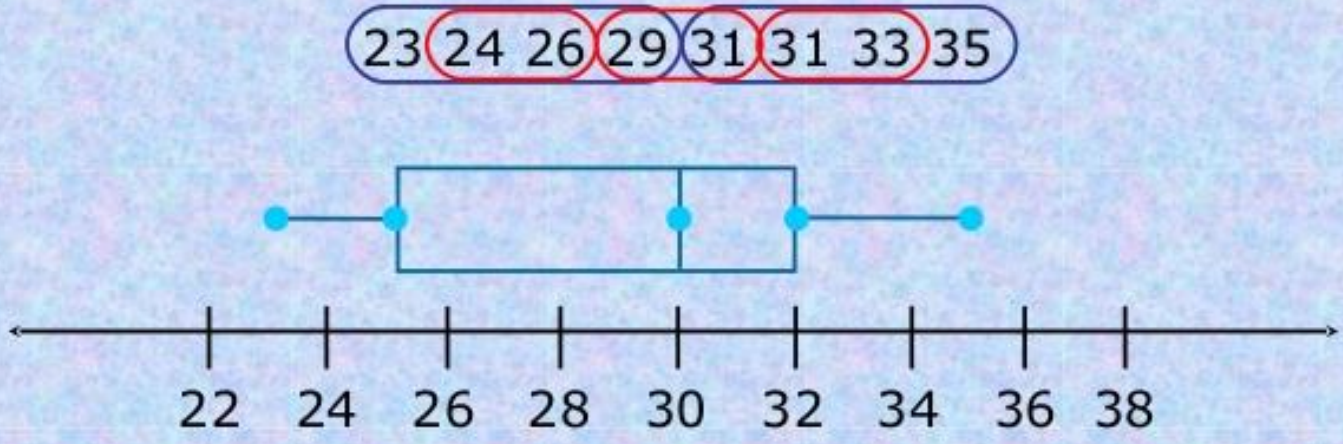


Side by Side Boxplots

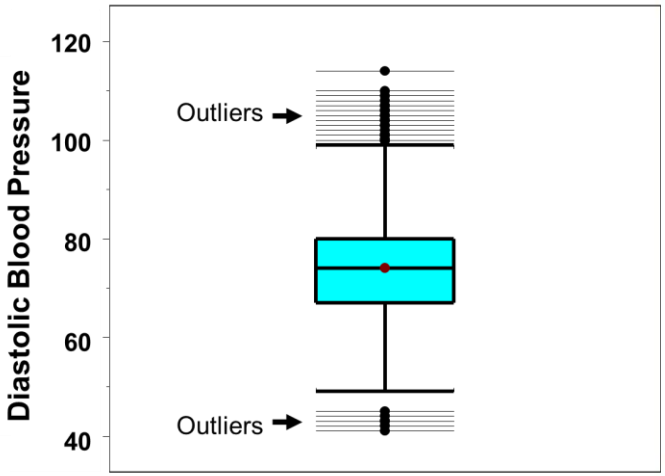


Data = 50, 60, 66, 70, 75, 80, 85, 89, 96
Minimum = 50
Lower Quartile = 66
Median = 75
Upper Quartile = 85
Maximum = 96

Side by Side Boxplots



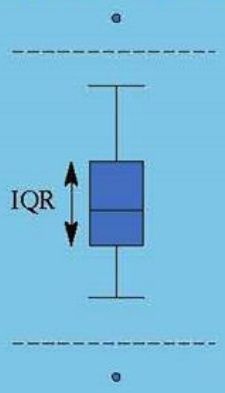
Side by Side Boxplots



Box plots with outliers

The box plot with outliers is a more sophisticated form of the box plot and is designed to identify any outliers that may be present in the data. How this is done is illustrated below.

Anatomy of a box plot with outliers



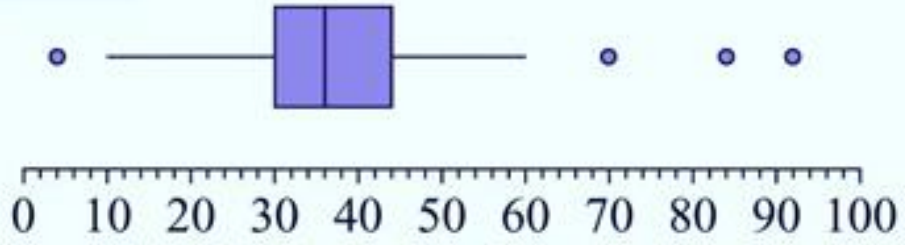
Maximum value:	possible outlier
Upper fence:	$Q_3 + 1.5 \times IQR$ (not drawn in)
Upper adjacent value:	highest data value inside fence
Third quartile:	Q_3
Median:	M
First quartile:	Q_1
Lower adjacent value:	lowest data value inside fence
Lower fence:	$Q_1 - 1.5 \times IQR$ (not drawn in)
Minimum value:	possible outlier

Two new things to note in a box plot with outliers are that:

- any points more than 1.5 *IQR*s away from the end of the box are classified as possible outliers (possible, in that it may be that they are just part of a distribution with a very long tail and we do not have enough data to pick up other values in the tail)
- the whiskers end at the highest and lowest data values that lie within 1.5 *IQR*s from the ends of the box

Box plots with outliers take more time to construct than standard box plots. However, they are normally constructed with the aid of a graphics calculator. Your prime task is to be able to recognise and interpret them, not just construct them.

Side by Side Boxplots



For the box plot above, write down the values of:

a the median

b the quartiles Q_1 and Q_3

c the interquartile range (IQR)

d the minimum and maximum values

e the values of any possible outliers

f the smallest value in the upper end of the data set that will be classified as an outlier

g the largest value in the lower end of the data set that will be classified as an outlier

Side by Side Boxplots

a median (vertical line in the box)	$M = 36$
b quartiles Q_1 and Q_3 (end points of box)	$Q_1 = 30,\ Q_3 = 44$
c interquartile range ($IQR = Q_3 - Q_1$)	$IQR = Q_3 - Q_1 = 44 - 30 = 14$
d minimum and maximum values (extremes)	$Min = 4,\ Max = 92$
e the values of any outliers (dots)	4, 78, 84 and 92
f upper fence (given by $Q_3 + 1.5 \times IQR$)	upper fence $= Q_3 + 1.5 \times IQR$ $= 44 + 1.5 \times 14 = 65$ Any value above 65 is an outlier.
g lower fence (given by $Q_1 - 1.5 \times IQR$)	lower fence $= Q_3 - 1.5 \times IQR$ $= 30 - 1.5 \times 14 = 9$ Any value below 9 is an outlier.